

説明可能なメラノーマ自動診断のための半教師あり学習およびマルチタスク学習を用いた診断指標の予測

著者	村林 誠也
出版者	法政大学大学院理工学研究科
雑誌名	法政大学大学院紀要．理工学・工学研究科編
巻	61
ページ	1-2
発行年	2020-03-24
URL	http://doi.org/10.15002/00022907

説明可能なメラノーマ自動診断のための 半教師あり学習およびマルチタスク学習を用いた 診断指標の予測

Towards Explainable Melanoma Diagnosis: Prediction of Clinical Indicators Using Semi-supervised Learning

村林誠也

Seiya MURABAYASHI

指導教員 彌富仁

法政大学大学院理工学研究科応用情報工学専攻修士課程

In this paper, we propose an effective method for predicting explainable melanoma indicators defined by a 7-point checklist in a situation where only a limited number of labeled data are available. Our proposal effectively utilizes virtual adversarial training as a semi-supervised learning framework with multi-task learning. This approach gives favorable performance for only a very limited number of expensive labeled data. The proposed method improves the final accuracy of melanoma diagnosis calculated based on these predicted indices by 7.5% (making it equivalent to expert dermatologists), based on 9,124 unlabeled images with diagnosis information added to the 226 base labeled training images.

Key Words : melanoma, computer-aided diagnosis (CAD), deep learning, semi-supervised learning, multi-task learning, 7-point checklist

1. はじめに

診断が難しい悪性度の高い皮膚がんであるメラノーマ（悪性黒色腫）の早期発見と高精度な診断実現のため、特に近年深層学習技術による自動診断手法の開発が行われている [1]。これらの手法は専門医に匹敵する高い識別精度を実現しているが、識別根拠を提示できず信頼性の面で改善の余地が残されていた。この問題に対し我々は臨床現場で利用される診断指標に基づくダーモスコピー特徴の数値予測器を試作した [2]。しかし、学習のために必要な教師データは医師に依頼する必要から作成コストが高く、極めて少量（226 例）のデータのみで構築された予測器においては、精度面での改善の必要性が残されていた。一方、少量のラベル付き学習データに加え、大量のラベルなしデータを学習に活用する半教師あり学習の枠組みの中で、Virtual adversarial training (VAT) [3] は、なめらかな識別境界を構築する制約を導入することで優れた成果を実現している。また、Multitask learning (MTL) [4] は、関連性の高い複数の 1 つ以上のタスクを同時に学習して共通の特徴表現を学習することで、汎化性能を向上する手法である。

本研究では VAT および MTL を応用することで、限られた数の教師つきデータと、それらの情報のないより多くの画像データを用いて上記診断指標の優れた予測器を構築した。本研究の目的は、他に多く見られる自動診断精度を追及する事ではなく、ブラックボックス化されたシ

ステムの結果に対する可読性・信頼性向上のために、診断根拠の定量的な提示を目指したものである。

2. 方法

（１）深層学習を用いた診断指標の数値予測器

本実験では、一般物体画像で事前学習済みの ResNet-101 を使い、臨床におけるメラノーマの診断指標である 7-point checklist [5] {S1:不規則な網構造, S2:青白い領域, S3:不規則な血管パターン, S4:不規則な枝状構造, S5:不規則な色素沈着, S6:不均一な点, S7:色素抜け構造}の有無[1, 0]とメラノーマ/母斑の計 8 項目の予測器を構築した。診断指標 7 項目の学習には 4 人の皮膚科専門医によって同項目への評価がなされた 226 症例のデータを用い、メラノーマ/母斑の学習にはメラノーマ/母斑ラベルのみ付与済みの 9,124 症例も用いた。

（２）半教師あり学習による準ラベルデータ活用

本実験では、半教師あり学習手法の一つである Virtual adversarial training (VAT) および Multitask learning (MTL) を参考に、教師データ付与済みの 226 症例に加え、診断指標の情報はないものの、最終的な診断結果であるメラノーマ or 母斑のラベルのみが付与された計 9,124 症例のデータも診断指標予測器の学習に利用し、予測能向上への影響を検証した。

表 1 7-point checklist 各項目の予測

ID	皮膚科医 (σ)	(i) Baseline (MAE)	(ii) VAT (MAE)	(iii) VAT+MTL (MAE)
S1	0.250	0.342	0.324	0.329
S2	0.209	0.200	0.141	0.154
S3	0.072	0.106	0.090	0.071
S4	0.154	0.244	0.327	0.369
S5	0.297	0.265	0.174	0.191
S6	0.250	0.314	0.299	0.322
S7	0.148	0.225	0.206	0.213
Ave.	0.197	0.246	0.223	0.236

表 2 7-point checklist に基づく診断結果†

評価指標	診断精度		
	感度[%]	特異度 [%]	AUC
皮膚科医	76.0	80.3	0.781
(i) Baseline	59.6	82.8	0.712
(ii) VAT	63.6	92.3	0.780
(iii) VAT+MTL	72.7	84.6	0.787

(3) 評価実験

研究では以下の 3 条件で予測器の学習を行い、比較を行った。

(i) Baseline : 診断指標ラベル付き 226 例のみ

(ii) VAT : VAT 適用, 226 例+9,124 症例

(メラノーマ or 母斑ラベル非活用)

(iii) VAT+MTL : VAT 適用, 226 例+9,124 症例

(メラノーマ or 母斑ラベル活用)

7-point checklist の各項目の数値予測能は平均絶対誤差 (MAE) で、7-point checklist の各数値から求められる診断結果については、10-fold cross validation 法で評価した。

3. 結果

7-point checklist 各項目の予測結果および皮膚科医 4 名による評価の分散 (SD) を表 1 に示す。また、7-point checklist に基づく診断結果の感度、特異度、area under the ROC curve (AUC) を表 2 に示す。

7-point checklist 各項目の予測において半教師あり学習適用後の予測誤差は 2.3% 程度改善した。また、診断指標の各数値から求められる診断精度は AUC で 7.5% 程度向上し、特にメラノーマ or 母斑ラベル利用時には皮膚科医を上回る結果を示した。

4. 考察

VAT の導入により、7-point checklist 各項に対する予

測誤差は Baseline に対し平均で約 2.3% 改善し、皮膚科医間の分散 SD である 0.197 に近い結果を示した。さらに、指標予測に基づく診断性能は、ROC で 6.8% 向上した。

また、VAT に加えて MTL を導入し、メラノーマ or 母斑ラベル情報を同時に学習することで、予測誤差の改善は約 1.0% にとどまったものの、指標予測に基づく診断性能は AUC で 7.5% と大きく向上し、診断精度はわずかに皮膚科医を上回った。以上から、我々の提案した手法は今回のような極めて少ないラベル付きのデータの学習において非常に有効であることが示された。

一方で、7-point checklist のような診断指標の予測に基づく診断は、同じ構成で最終的な診断精度だけを追求した識別器に対しては診断性能が下回り、人間の可読性に特化した診断指標の予測器には精度の面で限界があることが示された。我々は今回の実験で得られた知見に基づき、最終的な診断性能にフォーカスしたモデルおよび診断指標に基づく可読性の高いモデルを同時に用いたモデルの検討を進めることで、精度と信頼性を両立したシステムの開発を目指す。

5. 結論

本研究では、7-point checklist において定義されたダーモスコピー構造の予測器を作成し、極めて少量のデータに対し、半教師あり学習手法およびマルチタスクラーニングを用いることで異なるラベル情報が付与されたデータを有効活用し、臨床で使用される指標に基づく可読性高い診断において皮膚科専門医に匹敵する診断精度を示した。我々は、これらのダーモスコピー構造の定量化がメラノーマ自動診断の土台となることを信じている。

参考文献

- 1) A. Esteva, B. Kuprel, R. A. Novoa et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115-118, doi:10.1038/nature21056, Feb. 2017.
- 2) 村林誠也, 彌富仁: 深層学習器を用いた悪性黒色腫自動識別の識別根拠の評価. 第 37 回日本医用画像工学会大会, 2018
- 3) T. Miyato, S. Maeda, S. Ishii, et, al.: Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning. IEEE Trans Pattern Analysis and Machine Intelligence, 2018
- 4) R. Caruana, "Multitask learning," Machine Learning, vol. 28, pp. 41-75, 1997.
- 5) G. Argenziano, G. Fabbrocini, P. Carli et al.: Epiluminescence microscopy for the diagnosis of ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Archives of Dermatology No.134, pp.1536-1570 1998.